This is a working document prepared by the Energy Information Administration (EIA) in order to solicit advice and comment on statistical matter from the American Statistical Association Committee on Energy Statistics. This topic will be discussed at EIA's spring 2006, meeting with the Committee to be held April 6 and 7, 2006.

EIA-914 Data Expansion Challenges to Include Crude Oil Production

This paper will discuss the proposal to modify the Form EIA-914 to also collect crude oil production. It will also identify proposed modification of the EIA-914 natural gas production methodology which, if approved, would also be used to estimate oil production from the survey data It has two appendices, **Appendix A:** Form EIA-914, and **Appendix B**: Analysis of the natural gas data collected in the EIA-914 survey and the resulting production estimates.

1. The Proposal: Modify Form EIA-914 to also collect crude oil production data

Status

Domestic crude oil production accounts for 34 percent of total U.S. crude oil supply. Increased attention will be devoted to domestic supplies in light of the continuing high crude oil prices and renewed interest in reducing imports. At the same time, EIA needs a new crude oil production estimation system. EIA recently developed the Form EIA-914 system that has been very successful in producing natural gas production estimates. This system can be modified to collect crude oil production data and produce monthly crude oil production estimates.

Background

The Form EIA-914 survey collects data on natural gas production from selected largest States, the Federal Offshore Gulf of Mexico, and the aggregate production of the remaining lower-48 producing States. This survey data forms the basis of the Energy Information Administration's (EIA's) monthly estimates of natural gas gross withdrawals, marketed production, and dry production. Monthly natural gas production estimates are released 60 days after the end of the production month. The preliminary estimates of gross withdrawals of natural gas appear to be well within the 1 to 5 percent accuracy target range. In fact, the errors that result from sample design and modeling are less
than 1 percent.

EIA is currently at a high risk of not being able to produce timely and accurate crude oil production estimates. This is the case for weekly and monthly estimates at the State and national levels.

There are two major reasons that justify modification of the Form EIA-914 survey to collect crude oil and lease condensate production data:

> First, the Monthly Oil Production Update (MOPUP) system that is currently used has been in need of replacement for several years although budget constraints did not allow that to happen. MOPUP runs on an old PC loaded with legacy software that had to be taken off EIA's system for security reasons. Over a year ago the PC crashed and was given last rites by the Office of Information Technology. It was nevertheless resurrected by a Z Inc. contractor and has limped along since then.
>
> Second, MOPUP utilizes the monthly production data collected on Form EIA-182. The focus of that survey is to collect regional crude oil prices and it appears to have been satisfactory for that purpose. It never provided a stable base from which to estimate crude oil production albeit

that use of these data in the MOPUP system *as operated by experienced and expert staff* managed to produce good monthly estimates of crude oil production (including lease condensate).

Because the Form EIA-182 survey is going to be discontinued both a new crude oil production estimation system and a new survey instrument that captures monthly crude oil production data are needed.

Approach

EIA now collects two types of natural gas production data on its Form EIA-914 "Monthly natural gas production report", gross withdrawals and lease production. Part I of this form collects survey respondent information and Part II collects the production data broken out by Federal Offshore Gulf of Mexico, Louisiana, New Mexico, Oklahoma, Texas, Wyoming, and "Other States" (excludes Alaska). Thus the production data is collected at the State/area level for the six top producing areas.

A second report will be added to the Form EIA-914 family, the "Monthly Crude Oil and Lease Condensate Production Report". It will have a Part I with respondent information that will also collect two types of production data, crude oil and lease condensate. In order to have coverage for the six largest oil producing States the States of Alaska, California, and Colorado will be added to the list of areas used for gas production data collection. To cover the entire significant Federal Offshore domain the Pacific Federal Offshore (which has significant production and substantial resources) will also be added (see attached forms).

Respondent Burden

When the Form EIA-914 survey was approved by OMB, EIA had stated that the error estimates could be kept in the 1 to 5 percent range with an operator sample in the 250-350 range. Subsequent experience with the survey, improved modeling of estimate to estimate total production from the survey, and the fact that most large natural gas producers are also large crude oil producers make this possible for the crude oil production survey too. As illustrated in Figure 1, almost all wells produce both gas and liquids. If the dominant hydrocarbon is natural gas then the well is called a gas well and the produced liquid is called lease condensate. If the dominant hydrocarbon produced from a well is crude oil, then the associated gas is included in gross withdrawals of natural gas.

Survey System

The Standard Energy Processing System (STEPS) that is used to process the two types of natural gas production data collected on Form EIA-914 can be adapted to process the two types of liquids data (crude oil and lease condensate) that are proposed to be collected on the modified Form EIA-914.

From Survey Data to Crude Oil and Lease Condensate Production Estimates

The same formal methodology will be used as that used to make gas production estimates from the current Form EIA-914 survey data. A sample will be selected that covers over 80 percent of total

production. The production from the sample will be used to estimate the production from the thousands of operators that are not sampled.

Data Dissemination

The Reserves and Production Division (RPD), Office of Oil and Gas, will supply crude oil (including lease condensate) production data for publication in the Petroleum Supply Monthly, as it already does. Additionally RPD will maintain a web page that presents the methodology, the historical data files, and the separately reported current crude oil and lease condensate data and total production estimates.

Budget & OMB

Speaking of Challenges, the EIA-914 modification project has not been funded as of March 27, 2006, nor have we gone through the OMB clearance process, but we, of course, expect both to go expeditiously.

2. The Proposed Modifications to the EIA-914 Methodology

This section describes the proposed data estimation methodology used to estimate total production from respondent data. This will be a relatively qualitative presentation of this proposed methodology which focuses on the reduction of errors that result from assumptions in tested methodologies.

Gross Production Estimation for the Six Areas (Texas, Louisiana, Oklahoma, Wyoming, New Mexico, and Federal Gulf of Mexico)

A preliminary estimate of the final *Total Gross Production Rate* for each area is based on data provided by a cut-off sample of all operators for the data month. A cut-off sample was selected based on data for 2004.

Estimation

Gross Production Estimates for the Six Areas: A preliminary estimate of the final *Total Gross Production Rate* for each area (Texas, Louisiana, Oklahoma, Wyoming, New Mexico, and Federal Gulf of Mexico) is based on data provided by a cut-off sample of all operators for the data month. The preliminary total estimate is made each month by collecting gross production data from the sampled operators for the data month and adding to this an estimate of the gross production data from all operators *not* in the sample.

[1] $\hat{T}_t = S_t + \hat{N}_t$

This discussion will be focused on estimating the gross production each month,t, from all operators not in the sample, $\hat{N}_t$. A simple ratio model is given in equation [2] for any particular calibration year,c.

[2] $\hat{N}_t = (R_C) * (S_t)$

The value of R_C can assumed to be constant or variable over time. If assumed constant, it can be determined using variations of the classic Ratio Estimate Method for any area and time period for which the historical data are essentially complete. The ratio estimator, typically used for estimation with a cut-off sample, assumes that the sample coverage remains constant over time.

[3] $$R_C = \frac{N_C}{S_C}$$

As an example of this type of model, consider 2000 calibration year historical data:
Where

T_{00} = *Total Gross Production Rate in 2000=15,604 mmcf/day,*

S_{00} = *Sampled Operators Gross Production Rate* in *2000 = 13,658 mmcf/day,* and

N_{00} = *Not Sampled Operators Gross Production Rate* in 2000 = 1,945 mmcf/day.

Let

[4] $$R_{00} = \frac{N_{00}}{S_{00}} = \frac{1{,}945}{13{,}658} = 0.1424$$

For calibration year 2000, the model in equation [2] becomes

[5] $$\hat{N}_t = 0.1424 * (S_{T,t})$$

The estimate of $\hat{N}_t$, the Non-Sampled production, can be estimated from subsets of the total sampled production. The total Sample ratio model, along with a 1st Quartile model, the Upper 2 Quartile ratio model, and the Lower 2 Quartile ratio model are shown in Figure 3. The best performing constant ratio model was based on the Lower 2 Quartiles.

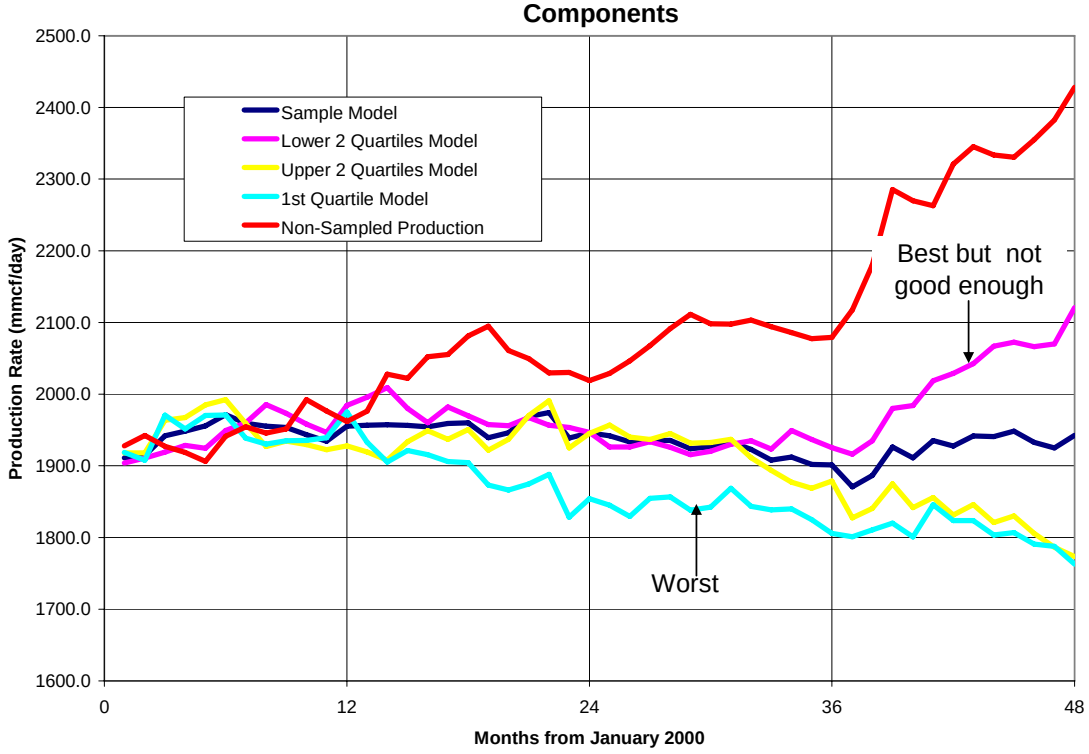
[6] $$\hat{N}_t = 0.2921 * (S_L)$$

The worst performing ratio model was based on the 1st Quartile of the sample production

[7] $$\hat{N}_t = 0.5210 * (S_F)$$

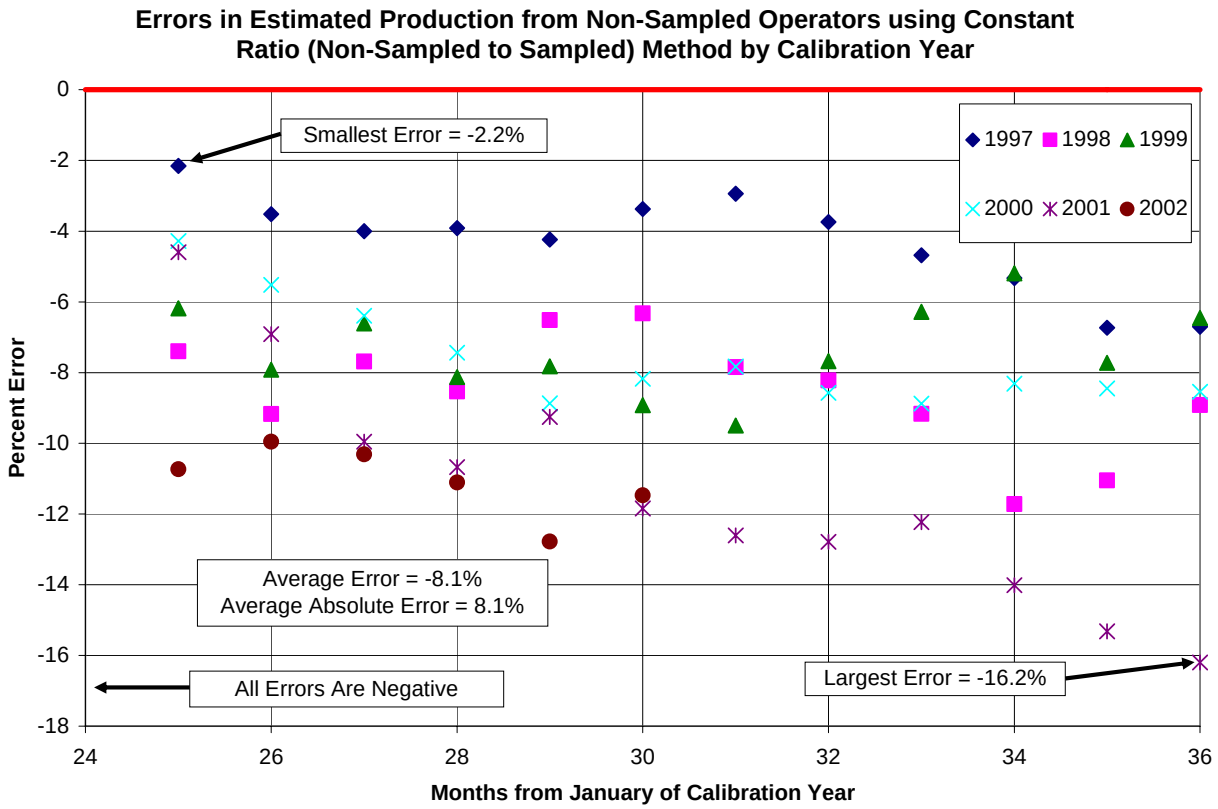
Similar results were obtained from the rest of the calibration years.

Figure 3. Constant Ratio Models of Non-Sampled Production by Sample Components

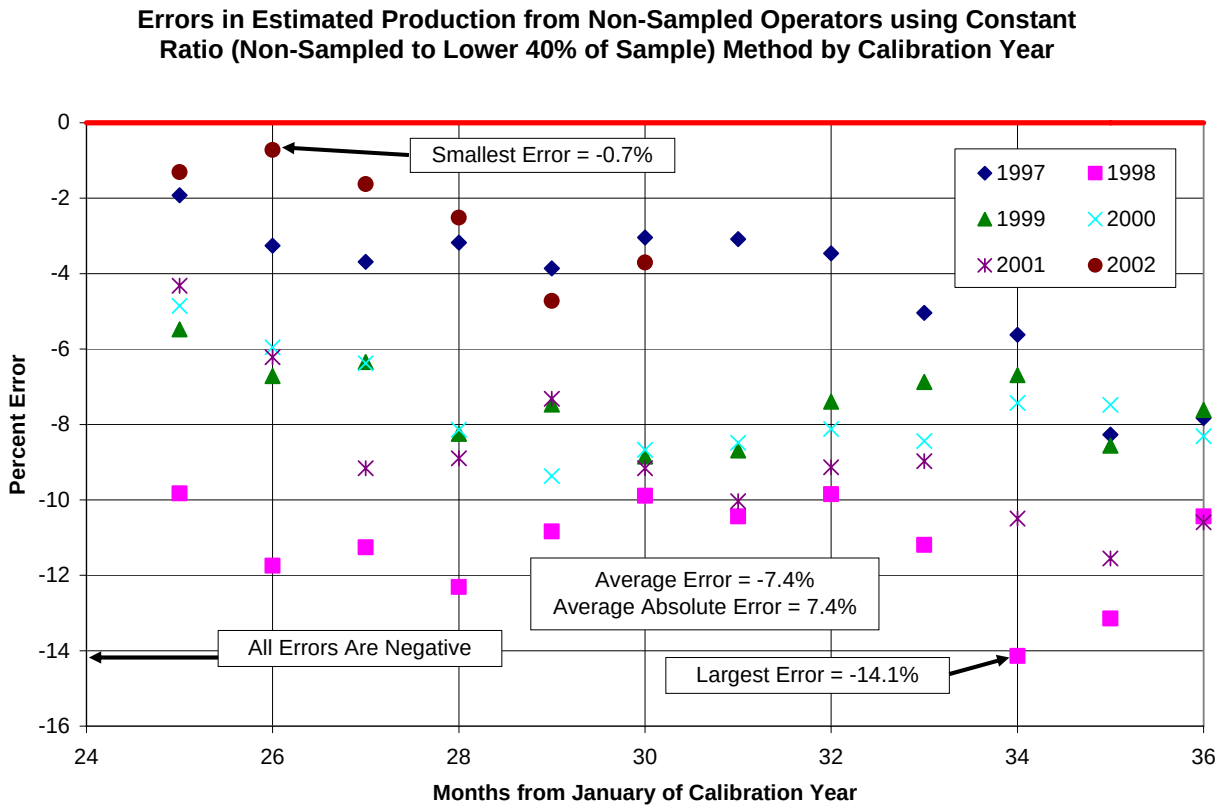


The preliminary total estimate will be made for each month in 2006 by collecting gross production data from the sampled operators for the data month, dividing by the number of days in a month to obtain an estimate for the gross production rate in billion cubic feet per day, and multiplying a subset of the sampled operators $S_{L,t}$ by a ratio.

The errors resulting from the various constant ratio methods were calculated.



The average absolute error and the largest error were somewhat lower when only the lowest e quartiles were used at 7.4 percent and minus 14.1 percent respectively.



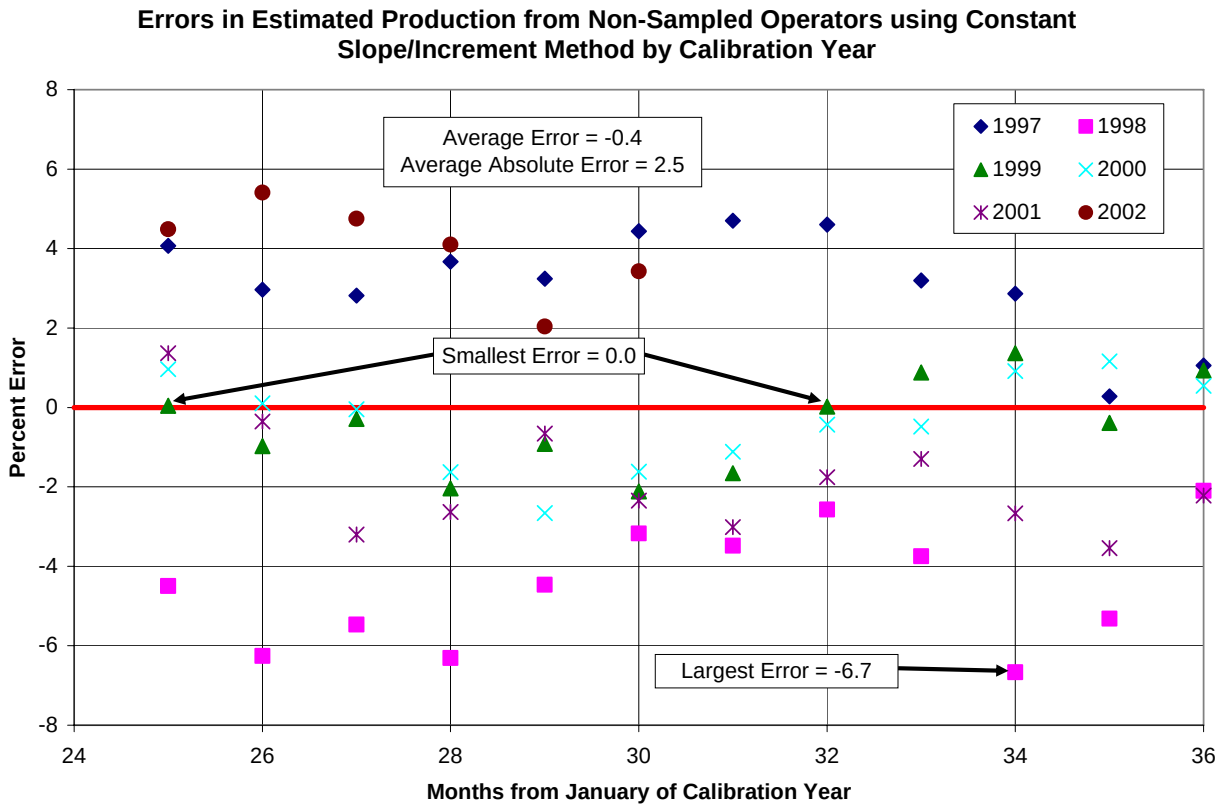
Models with variable ratios R_t were also tested.

[8] $\hat{N}_t = (\hat{R}_t) * (S_{L,t})$

These variable ratios had either constant or variable slopes. For the constant slope models,

[9] $R_t = (\hat{R} + \hat{a} * t) * (S_{L,t})$

where $\hat{R}$ and $\hat{a}$ are fit parameters. The errors associated with variable ratios were substantially smaller than those for constant ratio models. The average absolute error and the largest error were 2.5 percent and minus 6.7 percent respectively compared to the best constant ratio model errors of 7.4 percent and minus 14.1 percent respectively.



Somewhat better results were obtained using variable ratios that had variable slopes. Apparently the level of drilling for gas wells has a significant impact on the slopes. In equation **[10],** the D_t term depends on the level of drilling for natural gas at specific times.

[10] $$R_t = (\hat{R} + \hat{a} * t + \hat{b} * [D_t] * t) * (S_{L,t})$$

The average absolute error was 2.3 percent and the largest error in $\hat{N}_t$ was minus 5.6 percent. Remembering that $\hat{N}_t$ is less than 15 percent of the total production, the average absolute error in the estimated total production was less than 0.4 percent and the largest error in the six calibration years tested was less than one percent.

The sample selection and modeling will not lead to substantial errors. However, problems with survey data or basic calibration can lead to larger errors.

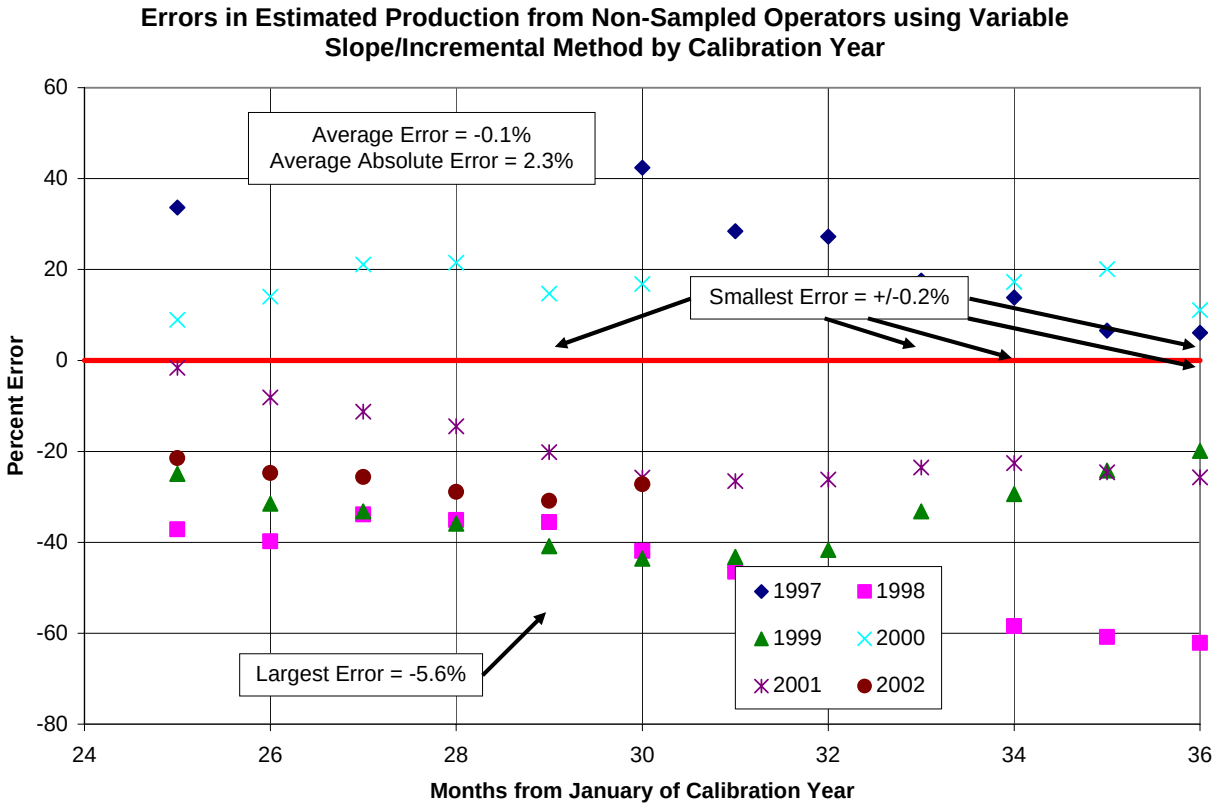


Figure 1
Figure 2